DOCUMENT RESUME

ED 256 153 FL 014 895

AUTHOR James, Gregory
TITLE Development of an Oral Proficiency Component in a Test in English for Academic Purposes.
PUB DATE 84
NOTE 36p.; Based on a paper presented at the Congress of the Association Internationale de Linguistique Apliquee (Brussels, Belgium, August 1984).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Communicative Competence (Languages); *English for Special Purposes; Foreign Students; Higher Education; Language Proficiency; *Language Tests; *Speech Skills; *Speech Tests; *Test Construction; Test Reliability; *Test Validity
IDENTIFIERS *English for Academic Purposes

ABSTRACT

The development of the oral component of the Test in English for Educational Purposes (TEEP) for foreign students in British universities was undertaken using information from previous oral test development efforts and a survey of 645 university admissions officers concerning students' communication skill needs. The resulting test is intended to provide an individual profile of the student's use of spoken English, particularly in academic contexts. Assessment is made with a test prerecorded on tape that the candidates listen to and respond to at intervals, ideally in a language laboratory. Much of the recorded information is also provided in printed form. Candidates demonstrate proficiency in free connected speech, oral summary of material presented graphically, asking and answering questions, responding to remarks and situations, and responding to and expressing opinions. The pilot test, which will be refined, uses five assessment criteria: appropriateness, adequacy of vocabulary for purpose, grammatical accuracy, intelligibility, fluency, and relevance and adequacy of content. An implicit objective of the test is to influence students' preparation for study in Britain and to enable them to become aware of their limitations and seek remediation. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

DEVELOPMENT OF AN ORAL PROFICIENCY COMPONENT IN A TEST IN ENGLISH FOR ACADEMIC PURPOSES

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Gregory James

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

GREGORY JAMES

University of Exeter

Paper presented at A.I.L.A. Brussels 1984 – expanded and revised. James

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Fragestellung

I don't care if a student speaks to me in sign language, as long as he's a good mathematician.
(Comment made by a lecturer during a survey of oral language needs of overseas students in British universities, May, 1983.)

A personal perspective

As an undergraduate, I spent a year as a student at the University of Geneva. Even with two years' university study of French at home behind me, I felt somewhat bewildered, if certainly exhilarated, at the prospect of having to construct a new rhythm of life in that language. From the hard-pressed Director of Studies, whom I consulted in my first week on a timetabling problem and from whose office I remember emerging even more confused than I had been when I had entered, so inadequately had I expressed myself in trying to explain my problem, and (to me) so rapidly and incomprehensibly had he barked his 'advice'; to the émigrée-Russian lecturer, who winced every time she detected an unFrench sound in her students' speech; from the sub-warden of the *cité universitaire* who used to take mischievous pleasure in insisting always on speaking in different varieties of slang and patois to foreign students; to the voluble and quick-tempered Ticinese student-restaurant manager whose French sounded as much Italian to me as mine must have sounded English to him – I had to get used to people going about their daily business through the medium of French, something my two undergraduate years (via the peregrinations of Saint Eulalia, the *Chanson de Roland* and *Polyeucte*) had ill equipped me to do. After the surprise and shock of a very bad result in the initial language test – at home I had always had good marks for language work! – I found myself in a class of two dozen or so non-native-speakers of French for the greater part of the day. To meet and converse with native-speakers on a regular basis was going to require a good deal of effort and will-power in that large, cosmopolitan city.

ED256153

FL014 895

BEST COPY AVAILABLE

Why a speaking test?

I have mentioned these apparently trivial personal details for the very reasons that, at the time, they were *not* trivial to *me*. My difficulties in oral communication were a source of some consternation; I remember frequently getting tongue-tied, I got elementary structural and morphological features of French wrong - I knew they were wrong as I uttered them, but in the pressure of communication I just could not help myself - and afterwards, I felt embarrassed at this loss of control. Sometimes, I just did not want to speak French at all.

My own experiences are minor, however, when compared with those of the overseas students from very different cultural, geographical, economic and linguistic backgrounds studying in Britain. The newness and unfamiliarity of the environment, together with initial language difficulties, can create in a student a crisis of confidence and a feeling of great insecurity, which certainly many can and do overcome, but which in some can lead to considerable personal distress, often, however, successfully dissembled from peers and teachers alike. It is the speaking skill that is the 'make or break' skill for students studying in a foreign language environment. Jordan & Mackay (1973:45), in one of the few studies of spoken English problems of overseas students, offer a sympathetic summary:

> One of the biggest problems that the overseas ..... student faces concerning the adequacy of his English, is his expectation of how quickly and easily he can improve it. Often by the standards of his own country his English seems good or at least adequate. It is only when he actually arrives in Britain that the full force of his language inadequacy strikes him. After arriving in Britain the overseas student has to cope with several factors simultaneously: strange food; finding accommodation; separation from family and possible homesickness; an alien culture and customs; settling into a new and often unfamiliar academic routine; general reorientation. All this plus language problems can cause the first few months to be very difficult. The student usually finds that it takes him much longer to settle into his new routine than he thought it would.

(See also Ryan, 1979 and the references given there.) Students may be quite unprepared for, and indeed unaware of, the effect their spoken language has on

BEST COPY AVAILABLE

native-speakers. Often, the fact of engaging another's attention through speech involves, on both sides, certain cultural assumptions which, in a foreign language situation, can rarely be made*. Jordan & Mackay's report concentrates on postgraduate students

> from foreign countries where they have studied through their mother tongue

and who, in the authors' opinion

> have considerable difficulty in studying in English when they first come to Britain
> (op. cit., 39)

In their investigation (of 106 students in 2 universities), they found that students estimated spending an average of 7 3/4 hours a week speaking to native-English speakers and 6 1/4 hours a week speaking English to non-native-speakers; 30% of respondents affirmed that, in general, they did not speak to university teachers; 12% had made no British friends in the survey average of 6 1/2 months' residence in Britain; 48% admitted that "speaking" (fluency, expression, discussion) was their biggest problem in English when they arrived at university. In their paper (*ibid*), Jordan & Mackay assert that students who

> come from Commonwealth countries and have English as their second or third language ..... will have had, on the whole, the bulk of their secondary and tertiary education through the medium of English. For the majority of these, English presents few problems .....

In my experience, however, it is just these students who, possibly confident in discussing their own specialist subjects, may have the greatest problems when interacting orally in English social contexts. They may have acquired considerable fluency in reading and writing, especially on academic themes, but they may be quite unused to oral communication with native-speakers of English. The kinds of cultural and rhetorical assumptions that they and their interlocutors would make speaking English in their home countries would not necessarily be those of their addressees in Britain. Unaware of their

BEST COPY AVAILABLE

inadequacies in British social English, they blunder through student society, often causing mirth, or at worst, contempt, and fail to understand the effect they are having on their native-speaker fellows, or why they are not 'accepted'. Not only do Jordan & Mackay fail to highlight this phenomenon, but also they tend to consider "speaking" as an all-embracing skill, irrespective of content and context, when making their analysis of students' problems.

Weir (1981; 1983b) in his survey of the difficulties experienced by students in studying through English at institutes of tertiary education in Britain, adumbrates the activities in which students have to speak English in their academic environment, noting (1981:146):

> oral interactive skills are not as important in terms of frequency of occurrence as the other three skills in the academic context.

This observation is borne out by the responses to our own survey conducted in May, 1983, amongst 645 tutors responsible for admissions in science and social science subjects in British universities.

How would you rate the importance for your course of an overseas student's ability in the following skill areas?

	H	M	L	N
Listening	579	63	1	2
Speaking	199	389	5	4
Reading	530	111	2	2
Writing	449	182	11	3

(H = very important; M = of medium importance; L = of low importance; N = no response. n = 645)

89.9% of respondents regarded *listening* as a *very important* skill for all students; 82.2% *reading*; 69.9% *writing*; and only 30.9% considered *speaking* as *very important*, with 60.3% rating it as *of medium importance*. We made no distinction in our survey between undergraduate and postgraduate admissions, and there was no way of knowing from the

BEST COPY AVAILABLE

responses which group respondents were referring to. However, given the numbers of overseas students in British universities, and their distribution amongst undergraduate and postgraduate courses (e.g., see Weir, 1983b:7-24 for the 1980-1981 statistics), it is significant that *listening* should emerge as the one skill which is considered *very important* by the majority of tutors. Many tutors are evidently extremely tolerant of students' efforts at expressing themselves - this may account for the lower rating of *speaking* in the survey - but they want to be sure that they are being understood by their students, both in lectures and in personal interaction. That they are not always understood is well attested in anecdotal evidence, and students sometimes appear extremely adept at concealing a misunderstanding or incomprehension of what is addressed to them. Within the social sphere, students will ostensibly have to operate very much as they do at home with friends, acquaintances and strangers, but within the socio-cultural conventions of English-in-Britain. Indeed, students, particularly those from outwith Europe, may be unaware of many of these conventions, and unwittingly cause perplexity or offence by not adhering to them. Students may hesitate to make a request or a complaint, for instance, because of their self-consciousness, their lack of confidence and their own perceptions of their inadequacies in pronunciation, command of idiom, syntax, vocabulary etc. Typical of students' comments on their difficulties in speaking English are (taken from Weir, 1983b:301-302):

> Generally my mood and state of mind play a significant role in my efforts to speak the language clearly and correctly.
>
> I think the main difficulty is that I cannot say what I really want to say.
>
> Some people cannot understand me unless I speak very slowly.

Echoes of my own experience as a student! Jordan & Mackay (1973:45) note:

> The student can only describe his language difficulties in a very vague and general way. He can say, for example, that he cannot understand English speakers speaking to each other. But when it comes to a careful analysis as to *why* he cannot understand, and *how* he can improve, he is naive.

BEST COPY AVAILABLE

Often, students are able to pinpoint quite accurately those areas of language in which they feel they are weak: pronunciation, vocabulary etc., but areas such as intonation, rhetorical convention or appropriateness are much more difficult to discern, let alone verbalise. Clearly there is a need for a procedure for the evaluation of proficiency in oral performance, which highlights areas of strengths and weaknesses and upon whose reported results realistic and informed decisions can be made as to the remediation considered necessary before a student can be considered to have an adequate command of spoken English to cope with the demands to be made on him/her in a course of study through the medium of English.

Towards an evaluation procedure for oral proficiency

In the Test in English for Educational Purposes (T.E.E.P.) developed by the Associated Examining Board (see Weir, 1981; 1983a; 1983b; this volume), the oral component comprises a taped test, administered along the lines of the well-established A.R.E.L.S. oral examinations in English as a foreign language (A.E.T., n.d.; A.R.E.L.S., 1976)[1]. A decision was taken by the A.E.B. at a very early stage in the development of the T.E.E.P. that, for administrative reasons, the oral component should, if possible, be a standard taped test, taken in a language laboratory. The literature abounds with arguments, both practical and theoretical, against the establishment of such a test format. McCall (1979:15), for example, writes:

> ..... the language laboratory, used in a teaching context, can instil fear and nervousness into students. ..... When the laboratory becomes part of the testing situation the fear and nervousness of the students are exacerbated.

No one would disagree that *any* unfamiliar setting for a test - particularly a test with the importance that the T.E.E.P. could have in a student's life - is bound to increase stress and apprehension. However, it has not been the case with the A.R.E.L.S. tests over the 20 years since their establishment that students have not been able, on the whole, to cope with the situation, if they

BEST COPY AVAILABLE

have been adequately briefed and trained. My own experience in preparing students for the A.R.E.L.S. examinations is that the vast majority actively enjoy the preparation, both for the type of tests they are and the challenges they offer. (Cf. Schulz, 1977.)

> The limitation of time and topic would not normally worry candidates, but an awareness of complete helplessness before a machine, coupled with the inability to go back and cover lost ground, can create hostility to the test device, if not the content matter of the examination. In order to minimise these problems students should be led to consider the examination as a simple extension of language-learning activities as practised throughout the course. It is absolutely essential that students receive training in the technique beforehand.
> (Beardsmore, 1972:16, describing the development of his own taped test procedure.)

Opinions differ on the effect a laboratory setting has on students. Whilst Bartz (1979:12) states

> ..... students view testing procedures requiring responses to be recorded not only as unrealistic and artificial, but also as highly threatening

Valdman & Moody (1979:556) claim that in their laboratory test for university students

> it seems that very few students visibly allow the unnatural setting to shake their ability to respond.

We have attempted, in the T.E.E.P.(O), to make the test as realistic as possible, within the evident constraints of the laboratory setting. The experience of the pilot tests leads us to conclude that students respond to this initiative very positively, and are prepared to 'play the game' with us. Candidates answered all the questions set, gave their comments, sometimes without prompt, and were able to deal with all the situations successfully. In many cases, students stated in the post-test questionnaires that they preferred a test in a language laboratory because they felt more secure than in a face-to-face situation. The common objection to laboratory tests, that they bear little resemblance to real-life communication situations and may be interpreted by students as artificial and irrelevant (cf. Taggart & Moore,

BEST COPY AVAILABLE

1972:27) has, in some measure, been met by the format we have adopted. In no way would we claim that the laboratory can replicate real-life situations, with the give-and-take of oral interaction face to face with others (for example, Ellis (1981:23) complains of the "sterility" of the A.R.E.L.S. tests); nevertheless, students have certainly been prepared, in our experience in the development of the T.E.E.P.(0), to make believe to the extent that they respond readily, even answering back to the tape. We feel that the format is less threatening than in many an interview situation - the commonest and most frequently reported oral test format - which can develop into a catechesis, often with examiners' marks reflecting their satisfaction with their own performances rather than those of the candidates. The positive comments in the questionnaires administered during the piloting of the T.E.E.P.(0) testify to students' favourable reactions to the test format and its content. (Negative comments tended to concentrate on aspects of procedure, e.g., alleged lack of time allowed for responses, rather than on the test *per se*)

> Evaluation of oral expression initially involves motivation and reality. The test situation must be motivating and real for the subject, so that he speaks freely and so that his production is representative of his general capacity to speak in most situations.
> (Robinson, 1971:260)

Apart from the A.R.E.L.S. examinations (for which see e.g., Fisher & Hubbard, 1979; Walker, 1981; 1983), the only other tests designed, administered and scored along similar lines to ours, which, as far as I am aware, have been reported in the literature, are Holeczek's (1972) secondary-school level test, Binham's (1976) Spoken English Grading Test at the Helsinki School of Economics and Valdman & Moody's (1979) Indiana University French Communicative Ability Test. (See Cartier, 1980 and Lowe & Clifford, 1980 for examples of other types of taped oral tests; cf. also Delattre, 1961; Schrand, 1970; Butzkamm, 1971; Beardsmore, 1972; Stoldt, 1973; Nübold, 1974; Folland & Robertson, 1976; Norrish, 1978; Thomaneck & Burgess, 1980; Berkoff, 1982.)

BEST COPY AVAILABLE

The Indiana test was developed by Valdman & Moody (1979) in response to their own criteria of the establishment of a valid measurement of communicative ability

> as opposed to pronunciation accuracy or spoken proficiency (op. cit., 554)

and the simulation of

> genuine conversational interaction in the target language (ibid.).

Ease of administration and reliability of scoring by "relatively inexperienced" teachers, were important considerations in the design of this four-part test, comprising

I Pictorially cued responses with six questions on tape;

II Personal questions;

III A question formulation exercise;

IV Situational responses.

(The question formulation exercise consisted of 3 sentences given, and candidates having to form an appropriate question to which each of the three sentences could be an answer; section IV was constructed very much along the lines of the A.R.E.L.S. model.) These components were scored along 3 parameters of assessment: 8 "points" were awarded for each section assessed, the criteria of assessment being:

- semantic pragmatic appropriateness 0 - 3 points;

  grammaticality and correct forms of lexical items 0 - 3 points;

  pronunciation accuracy and fluency 0 - 2 points.

BEST COPY AVAILABLE

In addition, Part IV included a score for "informational content".

> For example, for the situation, "You are in a café and you would like to order drinks for yourself and a friend" students might provide one of the following possible responses:
> (a) Garçon, apportez un café pour mademoiselle et une bière pour moi.
> (b) Mademoiselle prends une limonade et lui une tasse de thé.
> (c) Je voudrais du coca-cola et ce monsieur un jus orange [*sic*].
> In addition to pragmatic and semantic appropriateness, grammatical well-formedness and pronunciation accuracy and fluency, responses are judged on the basis of the amount of information communicated. For each situation, one point is awarded for the inclusion of one of a pre-determined number of informational elements. The situation given as illustration would require conveying a minimum of four informational elements: (1) term of address to waiter or waitress; (2) drink requested by speaker; (3) identification of other person; (4) drinks requested for other person. Thus response (a) would receive a score of "4" as would response (b), since the pronoun *lui* is sufficient in context; response (c), which lacks the term of address, would be scored only "3".
> (Valdman & Moody, 1979:556-557)

The point is contentious, however. A response such as

Deux bières, s'il vous plaît

would, apparently, be treated as an inadequate one in the context. The predetermination of informational content of response would appear to run counter to the prevailing acceptance of unpredictability in the evaluation of communicative discourse (see e.g., McDonough, 1979:84). Indeed, the authors themselves state (op. cit., 553):

> Valid evaluation of communicative ability requires a procedure that simulates use of language in an authentic situational context. The judgment must bear on the success of the communicative act and its semantic and pragmatic content rather than external well-formedness and accuracy.

Valdman & Moody (ibid.) criticise other test instruments, including the F.S.I. oral interview (see e.g., Lowe & Clifford, 1980), on the ground that

> on the basis of these tests it is difficult to separate a subject's communicative ability from such contaminating factors as general knowledge, inventiveness, intelligence, memory and imagination.

Are not these just some of the very factors that contribute to communicative ability, a nebulous concept inseparable from the personality attributes and

BEST COPY AVAILABLE

educational background of speakers? Whilst they take it as axiomatic that (*ibid*):

> Fundamental to the preparation of any instrument for the testing of communicative ability is the clear recognition of the lack of direct relationship between analytical knowledge of discrete linguistic elements (words, grammatical rules, phonemes, etc.), or the ability to manipulate these discrete elements in artificial exercises, and the ability to make use of these elements in engaging in meaningful communication interactions

their "communicative" laboratory test appears to fail in some measure to take account of this. They have delineated the informational content expected (whether this is made clear to candidates is not explicit), and deviation or omission is penalised, just in the same way that deviation from grammatical or other linguistic norms might be penalised in a more structurally oriented test; they have adjudged as "contaminating" just some of those features which contribute so effectively to a person's commmunicative ability: general knowledge, inventiveness, intelligence, memory and imagination. (Cf. Bachman, 1973:44-45.) Because these factors are so elusive and unquantifiable in the terms by which we, as language teachers, work, it is not uncommon to witness their avoidance by evaluators. What, in the end, is the difference betwen an artificial exercise set for evidence of syntactic and morphological manipulation, and one set for evidence of inclusion of meaning elements? Neither can be said to simulate any kind of "genuine conversational interaction" and the spurious veneer of "communication" in the responses to the situational contexts is unconvincing. We must ensure

> that in our procedures we are eliciting real communicative-linguistic behaviour and not artificial "test-speak".
> (Carroll, 1980:3)

Valdman & Moody's criteria of assessment, and the justfication for their definitions, seem ill-matched to the theoretical underpinning they have endeavoured to give their tests. Not only do we come across overlap in the criteria (e.g., *grammaticality and correct forms of lexical items*

BEST COPY AVAILABLE

appears to include pronunciation, which, ostensibly, is another category altogether; if a candidate makes 2 errors of pronunciation in a response, according to the descriptions given, he or she is penalised under the category of *grammaticality and correct forms of lexical items* as well as that of *pronunciation accuracy*, for the same error; only one error, on the other hand, would not incur a penalty under the first category, if, all other things being equal, the candidate merited a score of 3 in that category), but also basic contradictions among the theoretical dimensions of their test. Section III, the question formulation exercise, is no more than an artificially constructed drill, which, the authors themselves have proclaimed, bears little relation to the communicative ability they are purporting to measure. If, as they remark (op. cit., 555),

> While sensitivity to grammatical deviance varies from community to community, native [French] speakers do seem to expect a high level of correctness from learners who have acquired the language by formal training.

then that very sensitivity is part of the cultural environment of the language; to inculcate an awareness of grammatical accuracy and its importance in the socio-cultural life of the target language, it is not necessary to rely on the artificiality of drill exercises to elicit syntactically acceptable speech forms.

The authors do not indicate in their description of the test how the results are to be reported or interpreted, and they give no indication of whether they make explicit to candidates how weaknesses in certain areas might be strengthened, or how they would arrive at such decisions.

An earlier example of a taped oral test, administered in a language laboratory, is that reported by Holeczek (1972). He was concerned to establish, as objectively as possible, the communicative value of secondary school children's ability in a foreign language.

BEST COPY AVAILABLE

> Das Ziel des Fremdsprachenunterrichts auf der Sekundarstufe-I ist die aktive Beherrschung der fremden Sprache als Kommunikationsmittel ..... Die Frage, die sich auch für die Schulpraxis ergibt, ist, wie Lernfortschritte hinsichtlich dieser Zielsetzung objektiv überprüfbar gemacht werden können.
> (Holeczek, 1972:431)

The innovation of this test was that each section was assessed on different criteria, e.g., reading aloud a text - *intonation, pausing,* ..... ; retelling a picture story - *organisation of thought* ..... . We do not have any further details of this test, but we may note that the underlying theory and justification for it are reflected in the focussing of criteria of assessment that are discussed.

> Denn ..... solange ein Test nicht in der Lage ist, die erworbenen Fertigkeiten in der gesprochenen Sprache, die ja im Vordergrund unserer Bemühungen stehen soll, effektiv, zuverlässig und rationell zu messen, solange der Test letzten Endes nur ein Gegengewicht, eine Verbesserung gegenüber der bisher praktizierten Verfahren der schriftlichen Leistungsmessung ist, solange ist zwar im Schriftlichen - unter Aufgabe der Kontrolle der produktiven Fähigkeiten - eine gewisse Objektivierung erreicht worden, der prädominante mündliche Bereich dagegen weitgehend unberücksichtigt.
> (ibid., 432)

Butzkamm (1971) also offers a very brief description of tests with laboratory format, developed for first-year German students of English, with, e.g., situational responses and picture cues. Interestingly, in the criteria of assessment for his tests, he gives precedence to *meaning* over other aspects:

> Wichtig ist, daß Aussprache und Intonation keine Rolle spielen, solange die Äußerung verstehbar ist. Das erste Kriterium der semantischen Angemessenheit ist das wichtigste.
> (Butzkamm, 1971:8)

I have dwelt at some length on the formats of these reported tests because, although rarely cited in the literature, they acted as a preliminary informing source for our own test development; and secondly, because the discussion has highlighted some of the sorts of problems with which we have had to come to grips in that development.

BEST COPY AVAILABLE

Test design

Weir (1983a; 1983b; this volume) has detailed the theoretical background and *a priori* determination of the T.E.E.P. test specification. In our own survey, which supplements the research undertaken by Weir, we attempted to elicit from a sample of British university admissions tutors in a wide variety of science and social science disciplines, what information they felt would be helpful to them regarding an overseas applicant's spoken English. What emerged hints at a pan-disciplinary flavour of needs. Whilst different disciplines evidently have different emphases and entail different linguistic demands on students, we are to construct a general test in spoken English for all potential students, irrespective of discipline, and it is therefore incumbent upon us to investigate the overall pattern of requirements and perceptions by tutors and attempt to synthesise these into a coherent test which will respond to the needs of the majority, whilst not neglecting those of the minority. (For a brief discussion of such problems, particularly with reference to the interpretation of score/band information, see Cripwell, 1983.) The following, taken from our survey of 645 admissions tutors in British universities, indicates areas of overwhelming agreement:

	Percentage of respondents in agreement that the parameter should be tested:
Expressing attitudinal meaning	24.3%
Initiation in discourse	18.1%
Maintenance of discourse	25.6%
Termination of discourse	17.8%
Expressing information explicitly	88.4%
Expressing conceptual meaning	73%
Indicating the main point of an argument	63%
Summarising written text	62%
Planning and organisation of information	52%

The configuration of responses suggests that these admissions tutors are less concerned over the skills of interaction (initiation, maintenance and termination of discourse) than with those of the expression of surface meaning. This has, admittedly, made our task in the construction of a taped oral test much easier than it would have been if tutors had overwhelmingly

BEST COPY AVAILABLE

indicated a preference for a test of interaction skills; it may nevertheless demonstrate a general ignorance of the significance of these discoursal skills in interaction, and scepticism as to whether they can be isolated, taught or tested. We have, in the light of these responses and other pieces of research, concentrated on those areas which the majority of tutors appeared to agree upon as essential components of a test of spoken English, giving them the information they appear to want. At the same time, we did not overlook those areas which manifested less general agreement, but which were obviously areas of concern to a significant minority.

A further question in our survey requested tutors to rank in order of priority the 7 skill areas which they considered it would be most useful to have information on about a potential student's ability in spoken English. The pattern which emerged is one where the combined ranking of the skill areas suggested is as follows:

Skill area	Rank
Asking questions	1
Answering questions	2
Giving opinions	3
Explaining	4
Giving counter-arguments	5
Giving extempore reports	6
Reading aloud	7

In the initial stages of the T.E.E.P. oral test design, there was close co-operation between the A.R.E.L.S. Examinations Trust (A.E.T.) and the Associated Examining Board, and the format of the A.R.E.L.S. Certificate and Diploma examinations was taken as a starting point for the T.E.E.P.(0). Essentially, the A.R.E.L.S. examinations comprise the elements of listening comprehension and oral production; for the T.E.E.P., there is a separate unit for listening comprehension in an academic environment (see Weir, 1983b; A.E.B., 1984; and see Low, 1981:6 for a critique of the type of methodology adopted in the T.E.E.P. listening component), and it was found necessary to adapt the A.R.E.L.S. format, in particular the marking scheme, to reduce the

BEST COPY AVAILABLE

degree of reliance on listening comprehension in the production of speech. This may be quite unrealistic but given that the test is taped, taken in a language laboratory and that there are no facilities for candidates to request feedback, repetition etc., and no paralinguistic backup available to speech, we felt that total reliance on material heard through earphones would be unfair to the candidates, if not quite invalid for the test purpose. We feel that in the pilot tests we have achieved a constructive compromise.

The A.R.E.L.S. Certificate and Diploma examinations consist of six sections each, viz., (from A.E.T., n.d., 5-6):

1. Social responses
2. Intelligible speech
3. Audial comprehension
4. Sustained speaking
5 Oral accuracy
6. Free oral expression

Section 1 (Social responses)

> tests the candidate's ability to understand and be understood in everyday social situations: requests, invitations, thanks/apologies and reply, complaints etc. Immediacy of reply is important.

The A.R.E.L.S. examinations typically include brief remarks of the type:

> *You have just bought some things in a large shop, and you want to pay by cheque. You don't know the exact name of the shop to write on the cheque. How would you ask the shop assistant for it?*

Both the remarks and the situational descriptions are given orally only. As McCall (1979:18) has noted, in this section

> we have tested no more than the ability to complete a stimulus-response pair.

BEST COPY AVAILABLE

We decided, however, to retain this segment, because we considered that its inclusion would help candidates settle down at the beginning of the test. Moreover, it is in the appropriateness of remarks in just such stimulus-response contexts that overseas students will find readier acceptance from native speakers than if they falter into incoherence at the simplest greeting or conventional linguistic move. The T.E.E.P.(O) pilot version contained 9 remarks for quick response, each given once only and by a different person - 3 female and 6 male voices were used. Remarks were of the type:

1. *How are you?*
2. *Switch the light on, would you?*
3. *What was it you said your name was?*

6 seconds were allowed for each response. In only one case was there evidence of possible difficulty in comprehension or interpretation of response: the remark

It's hot in here

elicited such responses as:

No he isn't

No I don't think he is

No I don't think so.

It is impossible to tell whether these responses are genuine ones to the original remark (i.e., the personal pronoun is being used inappropriately to refer to atmospheric condition), or ones to a misinterpretation of the remark as a question: "Is *X* (in) here?". Items such as this have been discarded in the final test.

The second part of this section consisted of 13 situations including, amongst other things, requiring candidates to ask for specific information, make an

BEST COPY AVAILABLE

appointment, ask for permission to do something, make a request and a suggestion, arrange a meeting, give a warning and an apology; in no case did candidates have to play the rôle of another person, but they had to imagine that they were addressing a variety of different people, e.g., a tutor, a security officer, a receptionist, a fellow student. (Cf. Bonin & Birckbichler, 1975:22)

> An important ..... feature of oral tests is the problem of assessing the manner in which a speaker functions when faced with specific social variables which determine the status relation between him and the listener.
> (Trengrove, 1977:106)

Situations were presented as instructions of the type:

> *You arrive late for a class. Apologise to your tutor.*

In the A.R.E.L.S. examinations, these descriptions are given orally only. The the T.E.E.P.(O), they are also presented in printed form to minimise candidates' lack of comprehension of the situation envisaged. Response time varies - in the pilot test, from 5 to 12 seconds - depending on the situation given. The time allowed for responses throughout the tests was approximately 15% more than the average of that taken by a native-speaker control group in each case; at some points, the time allowed was increased slightly, after initial trialling with overseas students of differing abilities in oral production.

Section 2 (Intelligible speech) of the A.R.E.L.S. examinations tests candidates' intonation, pronunciation, stress etc. via a reading aloud exercise.

> Ability to deal with abbreviations, symbols, fractions, letters and spelling out, measurements, etc., and to read technical documents over the telephone, is assessed.
> (A.E.T., op.cit., 5)

Reading aloud as a test of oral proficiency has a considerable pedigree as a technique, but has long been a matter for debate, and the literature abounds

BEST COPY AVAILABLE

with more or less polemical outbursts on the subject. We felt that, in any case, if we were to retain this exercise, any text would have to be relevant to the academic context and approximate to an academic expository style, possibly including figures, fractions and/or abbreviations. The two major criteria for the retention of reading aloud as a test in an academic context would have to be: (i) it replicates a task which students have to perform regularly in an academic setting; and (ii) the information to be gained from the exercise would be significant with respect to the objectives of the design of an oral production assessment in English for academic purposes. The evidence for (i) is that students do *not* have to read aloud regularly; often when they do, it is material they themselves have prepared for presentation to their peer group (see Weir, 1983b:273-276). All the foci of assessment which one could muster for incorporation into a marking scheme for reading aloud can be covered in other parts of the test, and we decided not to adopt this section for the T.E.E.P.(O). However, the usefulness of the task of transferring written information to speech was recognised, and our adaptation of other sections of the A.R.E.L.S. examination format does incorporate this element with a more satisfactory scheme of assessment.

Section 3 (Audial comprehension) of the A.R.E.L.S. examinations tests the

> ability to understand spoken English in context and possibly adverse conditions

at the Certificate level, and at the Diploma level:

> the ability to understand ad lib, unscripted discussions on abstract topics in context ..... also ability to distinguish between the finest differences in voice inflexion and intonation.
> (A.E.T., ibid.)

Listening comprehension is tested explicitly elsewhere in the T.E.E.P., and in spite of the inevitability of a certain overlap in the natural dependence of oral production on a degree of aural comprehension, this is not formally

BEST COPY AVAILABLE

tested in the oral component. *Audial comprehension* is replaced by a section testing responses in context. The taped format precludes testing conversational ability *per se*; it cannot allow candidates to take the initiative or control and direct the course of the conversation. However, we felt that there was a good deal to be gained from reverting to the general pattern of an older model of A.R.E.L.S. examination format, in which candidates have to imagine they are in a small discussion group. Candidates listen to an edited, unscripted discussion, and at certain points they have to answer questions or give their opinions. In the pilot test, candidates had to give 8 such responses: an introduction, answers to 3 open-ended questions, and 4 opinions on various aspects of the discussion topic. The time allowed for responses varied between 10 and 30 seconds. In the discussion, which was an edited, unscripted, taped tutorial on a subject related to the theme of the reading, writing and listening components of the T.E.E.P., candidates were exposed to a variety of voices and regional accents of English, the natural speech of 3rd year science and social science undergraduates. The questions to be answered and requests for opinions to be given were so framed as to lead candidates to answer from their own experiences. The actual tutorial conversation formed a data base to which candidates might refer if they needed or wished to.

> If students are to communicate in the foreign language, we must recognize that they are more likely to want to communicate about something that concerns them in a very personal way. ... a student's motivation to speak is much greater if what he says is "for real" and if he is viewed as an individual whose opinions, feelings, and experiences are considered eminently worthy of attention.
> (Bonin & Birckbichler, 1975:22)

In this aspect of the test construction, the T.E.E.P. differs considerably from the older A.R.E.L.S. model, which was essentially a test of listening comprehension. (Cf. Berkoff's (1982) suggestion for an "interaction-type" format for a taped examination at *Bagrut* level in Israel, to replace the traditional interview-type oral examination:

BEST COPY AVAILABLE

> This would be a discussion between three people which would be taped. After the first minute or so one of the three will stop talking and whenever her or his turn comes the tape will be silent and the examinee will be expected to respond instead. The examiners will evaluate those responses. ..... The advantage of this method is that the examinee will be "directed" in his replies and thus cannot produce a prepared speech; it will prevent extraneous digressions; and it will free the class teacher from having to intervene to revive a dying conversation.
> (Berkoff, 1982:34)

He does not, however, make it explicit how he envisages this examination to be given, but it is evident that the format lends itself to language laboratory administration. We do believe, nevertheless, that the format and content construction of the T.E.E.P.(O) is superior to the *Bagrut* proposal put forward, because of the contextualisation of the tutorial and the incorporation of the candidate's own opinions within the framework of the conversation taped.)

Section 4 (Sustained speaking) of the A.R.E.L.S. examination requires candidates at the Certificate level

> to sustain speech along predetermined lines by telling the story of a sequence of pictures

and, at the Diploma level, to

> deliver[ ] a "subjective" summary (summary with personal comments) of a live interview or discussion .....
> (A.E.T., ibid.)

Neither of these tasks was considered of any relevance to the target population of the T.E.E.P., and we did not adopt them. However, the interpretation of graphic material, and specifically, requesting information on material which has not been fully understood does provide a task of direct relevance to students in an academic context. In the T.E.E.P.(O), we adopted the format of presenting graphic information, e.g., a bar or pie graph with certain data missing. Candidates have to identify the missing elements (the

BEST COPY AVAILABLE

location of which is indicated), and request clarification. Candidates areallowed a certain time to study the material, after hearing and reading the instructions on what to look for and say; in the pilot test the preparation time was 1 minute, response time 30 seconds. In this task, involving asking questions relating to non-verbal information, candidates who took the pilot tests encountered the greatest difficulty in expressing themselves, principally through lack of adequate vocabulary and through inability to frame coherent questions on the data presented. (Cf. McDonough, 1979.)

Section 5 (Oral accuracy) of the A.R.E.L.S. examination is

> a test in accuracy in the control of structures and words.
> (A.E.T., ibid.)

The potential usefulness of a section specifically to focus on oral accuracy is probably outweighed by the consideration of lack of face validity within the whole test context. We did not adopt this format for the T.E.E.P.(O). Despite our earlier observations concerning the breakdown of syntactic control under the pressure of communication, we have incorporated an element of assessment of grammatical accuracy in the T.E.E.P.(O) through other sections, and the marking scheme explicitly provides for evaluation of grammatical performance in certain areas.

Section 6 (Free oral expression) of the A.R.E.L.S. examinations tests whether candidates can, at the Certificate level

> show persuasiveness and clarity in speaking on a prepared subject of [their] choice

or, at the Diploma level,

> show how well [they] can present a brief discourse on a chosen theme with confidence, conviction, authority and in an interesting way.
> (A.E.T., ibid.)

BEST COPY AVAILABLE

We did not adopt this format, but replaced it with another data-based exercise, in which candidates were given a printed passage to study. They had 5 minutes to prepare an answer to a question set on this material, followed by 2 minutes in which to present their answer. In the pilot test, 2 passages (one of 411 words, the other of 447) were included, both on themes related to that of the tutorial discussion. Both texts contained a certain amount of statistical information, and for each, candidates had to answer a specific question designed to elicit an interpretation of the text and incorporation of relevant figures, dates etc. The aim of the task was to offer candidates the opportunity of reading a text, extracting and marshalling relevant information from it, and expressing this in their own words. The aim was not for candidates to read the text aloud, although, inevitably, there would be an element of this. During the pilot tests, an inordinate number of candidates read indiscriminately from the texts, and although the marking scheme was constructed in such a way as to penalise wholesale reading aloud, we concluded that little useful purpose would be served by retaining this section

in its original form. Consequently, subsequent tests will include a redesigned task, where little or no material will be presented via verbal text. Candidates will have to interpret information presented almost exclusively in graphic, rather than written, form.

The pilot test lasted 47 minutes, allowing a maximum of just over 11 minutes' candidate speech, including a one-minute self-introduction and the 2 two-minute text-based responses. Of the total test time, 16 minutes were devoted to reading and preparation time; the rest was listening to instructions and stimuli. The final version of the test will be considerably shorter, something under half an hour, eliciting a maximum of just over 9 minutes' candidate speech. Of the total test time, 11 minutes will be for preparation and reading: 5 minutes at the beginning, for reading through the test booklet briefly, and familiarising oneself with the requirements of the

BEST COPY AVAILABLE

test; one minute for the preparation of the graph/notes question; and 5 minutes for preparing a response to the question set on the graphic material. (In comparison, a British Council/E.L.T.S. interview can last as little as 7 minutes, eliciting something under three minutes' candidate speech, with no more than three sentences uttered at a time; I base these data on recordings made of live E.L.T.S. interviews in 1984.)

Summary

The aim of the T.E.E.P.(O) is to provide, by means of an individual profile, information on students' use of spoken English, particularly in academic contexts, with the specific objective being for the candidates to produce adequate spoken English for taking part in academic discussions and reading papers. The assessment is undertaken through a test pre-recorded on tape, which the candidates listen to and respond to at appropriate moments, ideally in a language laboratory. Much of the recorded information, including all the instructions, is also prov ded in printed form. Candidates may be expected to demonstrate proficiency in free connected speech, oral summary of material presented graphically, asking and answering questions, responding to remarks and situations and responding to and expressing opinions.

T.E.E.P. ORAL COMPONENT TEST FORMAT

Part 1	**Introduction**
Part 2	**Academic and social social responses**
Part 3	**Tutorial simulation**
Part 4	**Data search**
Part 5	**Data-based response**

BEST COPY AVAILABLE

Criteria of assessment

For the pilot test, we used 5 criteria of assessment, each on a 0 - 3 band scale, viz.,

Appropriateness	APP
Adequacy of vocabulary for purpose	VOC
Grammatical accuracy	GRA
Intelligibility	INT
Fluency	FLU
Relevance and adequacy of content	REL

In each band, there is a description of the level of proficiency. The highest band (3) represents a minimal competence, and is not equated to native speech. As examples, we give the band descriptions for APP and INT, as modified from the pilot test design.

Appropriateness

0 Unable to function in the spoken language; no response.

1 Able to operate only in a very limited capacity; responses characterised by socio-cultural inappropriateness.

2 Signs of developing attempts at response to rôle, setting etc., but misunderstandings may occasionally arise through inappropriateness, particularly of socio-cultural convention.

3 Few errors in the socio-cultural conventions of language; errors not significant enough to be likely to cause major social misunderstanding.

BEST COPY AVAILABLE

Intelligibility

0 Severe and constant rhythm / intonation / pronunciation problems cause almost complete unintelligibility; no response.

1 Strong interference from L1 in rhythm / intonation / pronunciation, or understanding is difficult, and achieved often after frequent marker repetition.

2 Rhythm / intonation / pronunciation require concentrated listening, but only occasional misunderstanding is caused or marker repetition required.

3 Articulation is reasonably comprehensible to native speakers; there may be a marked "foreign accent" but almost no misunderstanding is caused and marker repetition is required only infrequently.

Considerable further research is necessary to refine the definitions so that they reflect the information sought and can be matched against performance. Each Part is assessed by a different configuration of criteria, allowing examiners to focus on a variety of aspects of the candidates' spoken language (cf. Berkoff, 1982, for an example of an alternative schema):

Part 1 (Introduction)	VOC / GRA / FLU / (REL)
Part 2 (Remarks)	APP
(Situations)	APP / GRA
Part 3 (Tutorial)	FLU / REL / VOC
Part 4 (Data)	REL / VOC / GRA
Part 5 (Summary)	GRA / FLU / REL / (VOC)

BEST COPY AVAILABLE

The criteria in brackets are those which will be added in the final test marking scheme, the modification of which has been made in the light of experience in the pilot test development. INT does not appear in the lists, but is used for all Parts, as an impression; pronunciation is not considered for each individual response, but a separate overall band score is given for INT at the end of each Part, 1 - 5. The final band for the T.E.E.P.(O) will depend on an averaging out of the total bands for each Part. In the pilot tests, a 0 - 3 scale was used, with 3 representing minimal competence in the oral mode, i.e., a candidate who achieves a Band 3 overall may be considered, on the strength of actual sampled performance, not to have such difficulties in spoken English as might hamper him/her in a course of academic study in Britain *as far as this mode is concerned*; his/her spoken English should be able to cope with the demands made upon it during a course of study through the medium of English. (Cf. Harrison, 1979:11-14; Cohen & Olshtain, 1980.) It has been proposed by the A.E.B.'s English for Overseas Students Working Party that a 0 - 4 scale be adopted in the final test (A.E.B., 1984:x). Within the constraints of the format, the physical setting and the limitation of time, we have endeavoured to construct a test which samples candidate performance representative of proficiency in the skill of oral production, bearing always in mind the purpose of the exercise and the use that will be made of the test results.

> There is a grave danger in reducing students to mere *applicants* needing qualification ..... and forgetting that ..... their actual use of English will affect the range and quality of their *social* experience, their *work* and any ..... *education or training* they undertake. (Dixon, 1977:2) (Cf. also Dyson, 1972:24)

BEST COPY AVAILABLE

Some preliminary conclusions

> On constate ... que "l'expression orale" est loin d'être spontanément spontanée. ..... on sait ..... que la peur de l'oral de type scolaire ..... est particulièrement forte chez ceux auxquels a été inculqué l'image d'un oral-reflet de l'écrit et pour qui tout "écart" de langage est l'objet de reproches et non ..... susceptible d'être indulgemment interprétée comme une manifestation de désinvolture.
> (François, 1977:47)

One of the implicit, yet significant, aims of the T.E.E.P. is to influence students' preparation for their study in Britain, through a washback effect in the language classroom. For the oral component, it is hoped that students will be able to become aware of their limitations and seek remediation in order to aim at a successful command of spoken English for their purposes. The test encourages candidates to exercise their judgement, and concentrates on appropriateness and fluency of speech and relevance of subject matter rather than on syntactic accuracy alone. Candidates used to traditional scholastic oral interviews, with their focus on grammatical points and prepared "topics" (cf. James, 1983) will, it is hoped, be refreshed and feel less threatened by our approach. If we are able, through our test, to offer the opportunity to students of 'showing their paces' in a context and setting accepted by them as satisfactorily replicating potential tasks in their study situation in Britain - and the findings so far from both teachers and candidates demonstrate that the vast majority credit the test with face validity on this score - as well as giving tertiary-level admissions tutors the kind of information about students' language proficiency that they might find helpful when making judgements; if fewer students arrive in Britain inorate in English through the preparation for the challenges of the test, then we may be able to avow some measure of success which a traditional interview test often cannot, and make a stronger claim than Beardsmore made for his own taped test at university level that it was

> not invalidated by the problems that surround[ed] it if one accepts that at the very least it [did] ... no less than that which is attainable by the traditional interview situation.
> (Beardsmore, 1972:20)

BEST COPY AVAILABLE

Notes

* As an example, one may cite the fairly well attested difference between modern Japanese and modern Western rhetorical practices.

> Usually, the average Japanese lacks communication training and his communicative ability is not well looked up to. A Japanese who speaks his own language beautifully and fluently is not necessarily appreciated among the people. ..... We respect and regard a person's sincerity and faithfulness by his non-verbal behavior not by the spoken words. ..... A Japanese says "He understands me even if I don't explain everything" whilst an Occidental says "Let me explain fully so that you will understand me".
> (Kato, 1976:164)

Japanese, in Western societies, are sometimes regarded as shy and self-effacing, and this characteristic is, in turn, associated with people of indecision, lacking drive and personality. This effect of apparent indecisiveness is exaggerated by the cultivated hesitancy in Japanese rhetorical convention.

> *Ma* or *Ma'ai* in the Japanese language, which in mechanical terms is a delay in response or a break in the smooth flow of sound production, is a recognised virtue in thoughtful discourses. In America, whose linguistic culture is quite different from Japan's, speed and consecutive utterances of many sentences tend to be regarded as a virtue.
> (Kato, 1977:184)

(Cf. also Labarre, 1947; Nakajima, 1967; Morrison, 1972; Bazen, 1978:42; Branham, 1980:13-15; James, 1982:118-120; Hidasi, 1984. Sloan (1957) points out, however, that public *recitation* is an ancient and respected art in Japan.)

† The T.E.E.P. was formerly known as the T.E.A.P. (Test in English for Academic Purposes), and was reported as such by Weir (1983a, 1983b). Full details of the current test are available in a booklet, which includes specimen papers for the listening, reading and writing components, obtainable at £2.50 from The Publications Department, Associated Examining Board, Wellington House, Station Road, Aldershot GU11 4BQ, England. Details of the A.R.E.L.S. examinations can be obtained from The Director, A.R.E.L.S. Examinations Trust, 113 Banbury Road, Oxford OX2 6JX, England.

BEST COPY AVAILABLE

Acknowledgements

An earlier version of this paper, originally prepared for presentation at the 1984 TESOL Convention in Houston (my attendance at which was prevented by bureaucracy and penury) was given at a seminar at the Hong Kong Baptist College in April, 1984. A summary of the paper was given at the A.I.L.A. Congress in Brussels in August, 1984 and (with a further discussion of some of the results of the pilot tests and changes made in the test in consequence) at a staff seminar at the Chinese University of Hong Kong in November, 1984. Pilot tests and associated experiments have been carried out at the Bell Schools in Bath, Cambridge, Norwich and Saffron Walden; the International School, Exeter; the Universities of Århus (Fremmedsprogspædagogik Projektet), Exeter, Reading and Tel Aviv; the British Council English Teaching Institutes in Hong Kong and Singapore; and Tarawa Teachers' College, Kiribati. To all those responsible for helping me to gather data in the pilot tests, and to Andrew Cohen (Jerusalem), Allan Emmett (A.E.B., Aldershot), Graham Low (Hong Kong) and Sonya Sanders (Tel Aviv) for their constructive comments on this paper, my grateful thanks.

BEST COPY AVAILABLE

References

A.E.B. (1984) *Test in English for Educational Purposes*, Aldershot: Associated Examining Board

A.E.T. (n.d.) *ARELS Oral Examinations: rationale, development and methods*, London: A.R.E.L.S. Examinations Trust

A.R.E.L.S. (Association of Recognised English Language Schools) (1976) "ARELS and the ARELS oral examinations," *Zielsprache Englisch*, 1, 1 38-39

Bachman, L.F. (1973) "Testing oral production," *Bulletin of the English Language Center* (Bangkok), 3, 1 41=58

Bartz, W.H. (1979) *Testing Oral Communication in the Foreign Language Classroom*, Arlington, Va.: Center for Applied Linguistics (Language in Education: theory and practice, 2, 17)

Bazen, D. (1978) "The place of conversation tests in oral examinations," *English in Education*, 12, 2 39-50

Beardsmore, H.B. (1972) "Testing oral fluency," *Rapport d'Activités de l'Institut de Phonétique 1971-1972*, 6 Brussels: Université libre de Bruxelles, 10-22

Berkoff, N.A. (1982) "Testing oral proficiency - a suggested new approach," *English Teachers' Journal* (Israel), 27 33-36

Binham, P. (1976) "Spoken English grading test," *Kielikeskusuutisia*, 4 61-65

Bonin, T.M. & D.W. Birckbichler (1975) "Real communication through interview and conversation cards," *Modern Language Journal*, 59, 1/2 22-25

Branham, R.J. (1980) "Ineffability, creativity and communication competence," *Communication Quarterly*, 28, 3 11-21

Butzkamm, W. (1971) "Mündliche Tests im elementaren Fremdsprachenunterricht," *Englisch*, 6, 1 5-9

Carroll, B.J. (1980) "Measuring the communicative value of oral performance," MS.

BEST COPY AVAILABLE

Cartier, F. (1980) "Alternative methods of oral proficiency assessment," in Frith, 7-14

Cohen, A.D. & E. Olshtain (1980) "Developing a rating scale for testing functional speaking ability," paper presented at the 14th TESOL Convention, San Francisco, revised as "Developing a measure of sociocultural competence: the case of apology," *Language Learning*, 31, 1 113-134

Cripwell, K.R. (1983) " 'Specifications for the British Council English Language Testing Service' - B.J. Carroll (1978): a comment," in Hill, C.P. (Ed.), *Testing in ESOL*, University of London Institute of Education (Department of English for Speakers of Other Languages), Working Documents 5 82-91

Delattre, P. (1961) "Comment tester la facilité de parole dans un laboratoire de langue," *Le Français dans le Monde*, 3 36-38

Dixon, J. (1977) "An open examination of the assessment of English," *English in Education*, 11, 2 1-8

Dyson, A.P. (1972) *Oral Examining in French*, London: Modern Language Association

Ellis, R.J. (1981) "The assessment of oral skills," *Teaching English*, 14, 3 19-23

Fisher, M.S.S. & R. Hubbard (1979) "The reliability of R.S.A., F.C.E. and A.R.E.L.S. oral examinations," B.Phil.(Ed.) research paper, University of Birmingham, summarised in "To pass or not to pass - an investigation into oral examination procedures," *Spoken English*, 14, 3 109-116

Folland, D. & D. Robertson (1976) "Towards objectivity in group oral testing," *English Language Teaching Journal*, 30, 2 156-167

François, J.R. (1977) "Traits spécifiques d'oralité et pédagogie," *Pratiques*, 17, 3 31-52

Frith, J.R. (Ed.) (1980) *Measuring Spoken Language Proficiency*, Washington, D.C.: Georgetown University Press

BEST COPY AVAILABLE

Harrison, A. (1979) "Techniques for evaluating a learner's ability to apply threshold level proficiency to everyday communication," Strasbourg: Council of Europe Document No DECS/EES (79) 77

Hidasi, J. (1984) "'Japanese' versus 'European'," paper presented at the 7th World Congress of Applied Linguistics, Brussels

Holeczek, K. (1972) "Aufwertung und Objektivierung der Leistung im mündlichen Sprachunterricht," *Praxis des neusprachlichen Unterrichts*, 19, 4 431-434

James, G.C. (1982) "English as an acquired language," in Burniston, C., *Creative Oral Assessment*, Southport: English Speaking Board, 109-123

James, G.C. (1983) "Oral examining in a communicative framework," *Journal of English Language Teaching*, 18, 3 99-106

James, G.C. (1984) "An oral test component in English for academic purposes," paper given at a staff seminar, Chinese University of Hong Kong, November, 1984

Jordan, R.R. & R. Mackay (1973) "A survey of the spoken English problems of overseas postgraduate students at the Universities of Manchester and Newcastle-upon-Tyne," *Journal of the Durham and Newcastle Institutes of Education*, 25, 125 39-46

Kato, H. (1977) "Some thoughts on oral examinations for advanced students in Japanese," *System*, 5, 3 181-186

Kato, S. (1976) "A method of evaluating oral communication," *Speech Education*, 4 161-167

Labarre, W. (1947) "The cultural basis of emotions and gestures," *Journal of Personality*, 16, 1 49-68

Low, G.D. (1981) "Communicative testing as an optimistic activity," *Working Papers in Linguistics and Language Teaching* (University of Hong Kong Language Centre - Special issue on language testing) 4 1-15

Lowe, P. & R.T. Clifford (1980) "Developing an indirect measure of overall oral proficiency," in Frith, 31-39

BEST COPY AVAILABLE

McCall, D.K. (1979) "The development of an oral test designed to assess foreign learners' ability to communicate and interact with a native speaker of English," M.Phil., University of York

McDonough, S. (1979) "Testing communicative competence in English as a foreign language with students of natural and social sciences," in Culhane, P.T. & J. White (Eds), *Testing*. University of Essex (Department of Language and Linguistics) Occasional Paper 21, 83-96

Morrison, J.L. (1972) "The absence of rhetorical tradition in Japanese culture," *Western Speech*, 36, 2 89-102

Nakajima, F. (1967) "The Japanese attitude toward spoken language," *English Language Education Council Publications*, 8 48-53

Norrish, N. (1978) "Using the language laboratory to test linguistic performance and communicative ability," *System*, 1 30-40

Nübold, P. (1974) "Sprechfertigkeitstest: Lernzielgerechte Leistungskontrolle im Sprachlabor und ihre technologischen Probleme," *System*, 2, 3 10-20

Robinson, P. (1971) "Oral expression tests," *English Language Teaching*, 25, 2 151-155 & 25, 3 260-266

Ryan, M. (1979) "Testing and assessment procedures for diagnosing proficiency in E.F.L. for use in science and technology," M.A., University of London

Schrand, H. (1970) "Zur Kontrolle mündlicher Leistungen im Sprachlabor," *Das Sprachlabor und der audiovisuelle Unterricht*, 1 2-8

Schulz, R. (1977) "Discrete-point versus simulated communication testing in foreign languages," *Modern Language Journal*, 61, 3 94-101

Sloan, T.O. (1957) "Public recitation in Japan," *Quarterly Journal of Speech*, 43, 4 394-398

Stoldt, P.H. (1973) "Kontrolle mündlicher Sprachfertigkeiten in Unterricht und Prüfung," *Praxis des neusprachlichen Unterrichts*, 20, 1 33-42

 BEST COPY AVAILABLE

Taggart, G. & G.A.B. Moore (1972) "The use of closed-circuit television for oral tests in the language laboratory," *N.A.L.L.D. Journal*, 46, 1 27-30

Thomaneck, J.K.A. & G.J.A. Burgess (1980) "Testing in the language laboratory - an argument for its efficiency," *British Journal of Language Teaching*, 18, 1 35-39

Trengrove, W. (1977) "Oral proficiency in English as a second language - the designing of scholastic tests," D.Ed., Stellenbosch University

Valdman, A. & M. Moody (1979) "Testing communicative ability," *French Review*, 52, 4 552-561

Walker, C.A. (1981) "Reliability and validity of global versus analytic mark schemes in oral proficiency testing," M.A., University of York

Walker, C.A. (1983) "The whole is more than the sum of the parts: evaluation criteria in oral tests," *British Journal of Language Teaching*, 21, 1 41-44

Weir, C.J. (1981) "What to test?" Report prepared for the English for Overseas Students Working Party, Aldershot: Associated Examining Board

Weir, C.J. (1983a) "The Associated Examining Board's Test in English for Academic Purposes: an exercise in content validity," in Hughes, A. & D. Porter (Eds), *Current Developments in Language Testing*, London: Academic Press, 147-153

Weir, C.J. (1983b) "Identifying the language problems of overseas students in tertiary education in the United Kingdom," Ph.D., University of London

***** *****

BEST COPY AVAILABLE